



**Ermittlung von Indikatoren für die
Kompetenzeinschätzung von Schülertexten
mittels computerlinguistischer Methoden
und dialogische Entwicklung eines
Prototyps für die computergestützte
Analyse von Aufsätzen**

Jasmine Bennöhr

Berlin, 16. Oktober 2009



Überblick

- Beispiele
- Motivation
- Ziele im Überblick
- Indikatorenermittlung - Methoden
- Onlinewerkzeug
- Korpuserstellung

Beispiele aus KESS4

- 1) *Ihr gibt den schuld für mich*
- 2) *Moritz versuchte hastich*
- 3) *auf einen großen Hügel des Dorfes*
- 4) *Ich mache nie wieder so ein Dramer um so einen blöden Koffer*
- 5) *Ich wünschte ich wüsste was das für ein Koffer ist*
- 6) *Am nächsten Tag wusste er was er sich wünschen würde*
- 7) *ein Plan der Hamburger Kanalisation, der in einer Schatulle eingeschlossen war*
- 8) *krikten 10 Euro von den Man den das Koffer gehorte*

Beispiele aus KESS4

- 9) *Die Kinder musen den Koffer zur Pullizt dring ader die Kinder haden gesat*

- Aufsatzbewertung ist komplex
- Wird Gesamtnote verteilt, besteht diese immer aus mehreren Dimensionen
 - Wortschatz
 - Grammatik (Satzebene, Wortebene)
 - Rechtschreibung
 - Ausführlichkeit
 - Registerwahl
 - ...
- Gewichtung, Begründung für Gewichtung
- Noten sind immer relativ, ein Vergleich über Klassenstufen hinweg ist nicht möglich

5

Kompetenzeinstufung - Kompetenzraster

Kompetenzstufe nach dem GER	Beschreibung der Kompetenz	Beispiele
C2	Durchgängig richtige Schreibung komplexer Wörter mit schwieriger morphematischer Struktur; Großschreibung in Wendungen, sichere Zusammen- und Getrennschreibung wird beherrscht (ausgewogenes Strategieprofil).	
C1	Richtige Schreibung komplexer Wörter sowie in komplexeren Satzstrukturen.	Kondensstreifen, widerspiegelten
B2	Richtige Schreibung auch morphematisch schwer zu durchschauender Wörter, angereichert durch einzelne komplexere Wörter bzw. Fach- oder Fremdwörter. Unsicherheiten in Wendungen, die zwar im Sprachgebrauch sind, aber vom Schreiber noch selten genutzt werden (wortübergreifende Strategie).	beabsichtigte; vorgenommen, funktionieren;
B1	Ein einfacher Wortschatz, angereichert durch einzelne komplexere Wörter, wird orthografisch richtig geschrieben. Unsicherheiten bei morphematischer Durchgliederung der Wörter (morphematische Strategie).	Hund, stärker
A2	Richtige Schreibung einfach strukturierter Wörter des Alltagswortschatzes (orthografische Strategie); kaum Fehler in komplexer gebildeten Wörtern.	Liebe, Mutter, Katze
A1	Verschriftungen der Artikulation (phonetisches Schreiben, alphabetische Strategie). Auslassungen, Verwechslungen von Buchstaben, Übergeneralisierungen.	schprang der Kofer auf
0	Rudimentäre Schreibungen, die diffus und unlesbar erscheinen	Hate ne ille si heuben beslosen

- Was macht einen guten Text aus? Gibt es Indikatoren? Welche? Lassen sich schon früh Merkmale in Texten finden, die die Wahrscheinlichkeit späteren Lernerfolgs/Lernmisserfolgs erhöhen?
- Leistungskontrolle: Wie können Aufsatznoten/Kompetenzeinstufungen gerechter vergeben werden? Einhaltung der Testgütekriterien, Explizierung von Bewertungskriterien
- Wie kann man Schüler beim Schreiben/Schreiblernprozess unterstützen (ohne Mehrarbeit für Lehrer) Selbstlernen: Autonomie der Schüler unterstützen
- Wann lernen Schüler welche linguistischen Merkmale? Gibt es eine typische Entwicklung?

7

- 1. Indikatorenermittlung und -untersuchung
 - Ermittlung von Indikatoren für die Qualität von Schülertexten
 - Analyse der Indikatoren für Textkompetenz im Längsschnitt
- 2. Praktische Nutzung der Ergebnisse
 - Prototyp eines Onlinewerkzeugs, das Lehrern und Schülern Rückmeldungen zum Text gibt
- 3. Aufbereitung der Analyseergebnisse für die theoretische Nutzung der Ergebnisse
 - Aufbereitung zur Weitergabe an Wissenschaftler, die Kompetenzmodelle (weiter-)entwickeln

8

- leicht automatisch zu erfassende Oberflächenmerkmale^[1], linguistische Merkmale, die relativ leicht maschinell zu erfassen sind^[2]

[1] z. B. Anzahlen für Zeichen (Wortlänge), Wörter, Sätze, Absätze, Satzzeichen (Satzende, ‚Kommata‘, Doppelpunkt, Ausrufezeichen), Großbuchstaben, Kleinbuchstaben, Umlaute, Worthäufigkeiten (Types, Zeichenketten, ohne Zurückführung auf Lexeme, Minimal-, Maximalwerte, Durchschnittswerte bezogen auf verschiedene Einheiten wie Wörter pro Satz, Sätze pro Absatz, Buchstaben/Zeichen pro Wort. [2] Wortarten (geschlossene Klassen: Konjunktion, Artikel, Personalpronomina, Präpositionen), Worthäufigkeiten (types, Lexembildung), falsche Präteritumformen/Vergangenheitsformen (rufte statt rief etc.), Substantivierungen (-heit, -keit, -ung), Silben (ver-, ab- ...), Zahlen, Wortfelder zu bestimmten Themen (eigene Meinung äußern, wo im Text, Anfang/Ende), Kookkurrenzen (Bigramme etc.)
- Text-Readability-Measures Lesbarkeitsformeln
 - Dale/Chall, Flesh reading ease, Flesh-Kincaid, SMOG, FOG, Bennöhr, Si/Callan concept difficulty
 - Satzlänge, Wortlänge, lexikalische Informationen, Kohärenz
- Statistische Verfahren
 - Latent semantic analysis / indexing (LSA / LSI) / Support Vector Machines (SVM)
 - Skalierungsverfahren
- Multidimensionale Ansätze
 - Quantifizierung linguistischer Merkmale + Faktorenanalyse (Biber 2009)
- Module von Programmen zur automatischen Textbewertung

Lehrerurteil versus Software

	Lehrerurteil	Software
Objektivität	Variiert je nach Einstellung zum Schüler (Sympathie, Erwartung), Laune, Tageszeit, persönlicher Einstellung zu geäußerten Meinungen	Immer neutrale Einstellung zum Text und Autor
Reliabilität	Bewertungsmaßstäbe variieren, da sie nicht explizit sind, Unterschiede bei wiederholten Bewertungen	Bewertungsmaßstäbe konstant und explizit
Validität	Misst ggf. etwas anderes als beabsichtigt unklar, was eigentlich gemessen wird und warum	Was gemessen wird, ist sehr gut steuerbar, ABER Komplexität ist eingeschränkt
Vergleichswerte, Normen	Einordnung der Leistung aufgrund der persönlichen Erfahrung in wenigen oder nur einer Schule	Einordnung der Leistung ins Gesamtspektrum einer großen Stichprobe z. B. eines Jahrgangs

Automatische Textbewertung

Name / Autoren	Methoden, Variablen und Besonderheiten
Project Essay Grade (PEG)	linguistische Oberflächenmerkmale: Aufsatzlänge (Anzahl der Wörter), Komplexität der Satzstruktur vertreten durch die Indikatoren Anzahl Präpositionen, Relativpronomina, weitere Wortarten, Variation in der Wortlänge, ohne Natural Language Processing-Technik und ohne lexikalische Informationen
Intelligent Essay Assessor	Latent Semantic Analysis (LSA), ohne Beachtung der Wortstellung im Satz, große Datenmengen nötig, aufwändige Berechnungen
Educational Testing Service I (ETS)	lexikalisch-semantische Methoden, manuelle Bearbeitung der Aufsätze nötig
Electronic Essay Rater (E-Rater) vom ETS	Kombination aus statistischen und symbolischen Methoden Diskursstruktur, syntaktische Strukturen, Wortschatzgebrauch Trainiert auf einem Korpus von 270 Texten sehr akzeptable Ergebnisse für die englische Sprache
Conceptual Rater (C-Rater)	logische Relationen
Bayesian Essay Test Scoring sYstem (BETSY)	Bernoulli Modell, Wortschatz
Intelligent Essay Marking Systems (IEMS)	Pattern Indexing Neural Network, Cluster
Automark	Standardisierung von Rechtschreibung und Zeichensetzung Identifikation von syntaktischen Konstituenten Templates für Inhaltsbewertung
Schema Extract Analyse and Report (SEAR)	Stil und Inhalt, Template
Paperless School free-text Marking Engine (PS-ME)	Mastertexte und 'negative' Mastertexte keine Echtzeitberechnungen
Markit (Williams und Dreher, 2004)	Nur eine Modellantwort nötig, Wörterbuch als Worddokument
Verfahren von Lenhard, Bayer, Hoffmann und Schneider (2007)	LSA - nur sehr kleine Stichproben

11

Rückmeldungen für Lehrer und Schüler

■ Lehrer

- 1. Der Aufsatz enthält überdurchschnittlich viele unterschiedliche Wörter im Vergleich zu anderen Aufsätzen von Schülern aus der Klassenstufe 4 (50 vs. Mittelwert 35). Der Schüler verwendet eine große Bandbreite an weniger gebräuchlichem Wortschatz (z. B. Kanalisation). Das deutet auf einen reichhaltigen Wortschatz des Schülers hin. Auf einer Skala von 0 bis 100 für die Klassenstufe 4 liegt die automatische Bewertung dieses Aufsatzes im Bereich Wortschatz bei 80.
- 2. Der Schüler verwendet nur einfache Wörter aus dem Basiswortschatz wie ‚Kinder‘, ‚Baum‘. [...]

■ Schüler

- In deinem Text hast du einfache Wörter verwendet. Lies einen Text in deinem Sprachbuch und suche Wörter, die deine Meinung ausdrücken wie z. B. ich „finde“ und benutze sie das nächste Mal in deinem Aufsatz.

12

■ KESS-Daten

Aufsatzdaten	N digitalisiert [1]	N bewertet [2]	Beschreibung der Testergebnisse
Klassen 2 / 3 - (Der geheimnisvolle Koffer - Das fliegende Kind - Katze u. Vogel)		Kl. 2: 13 - 12 - 65 Kl. 3: 34 - 34	
Klasse 3 - Sommercamp 2007 (1 Thema)	337	337	Grammatik, C-Test, KFT
KESS4 - 2003 (1 Thema)	ca. 8000	ca. 8000	KFT
KESS7 - 2006 (2 Themen)	ca. 1500	60	Leseverständnistest
KESS8 Vorerprobung 2007 (5 Aufsatzthemen)	ca. 100	ca. 100	C-, Grammatik-, Wortschatztests, Rechtschreibung
KESS8 - 2007 (13 Themen)	Mindestens 150 pro Thema	ca. 150 pro Thema	C-, Grammatik-, Wortschatztests, Rechtschreibung, Leseverständnistest
KESS10 - 2009 (voraussichtl. 6-10 Themen)	liegt noch nicht vor, ca. 250 pro Thema	liegt noch nicht vor, ca. 250 pro Thema	C-Tests, Rechtschreibung, Leseverständnistest

[1] mit Vergleichsergebnissen wie Sprachtestergebnissen etc. aus mindestens einer, meist mehreren Klassenstufen

[2] Aufsatzbewertung in bis zu 50 Bewertungskategorien, ggf. weitere Aufsätze und Kategorien nachbewerten lassen.

- **Genauere Sichtung des Materials**
- **Nacherhebung unterrepräsentierter Klassenstufen**
- **Nachdigitalisierung fehlender Texte**
- **Tokenisierung**
- **Lemmatisierung**
- **Festlegen von Tagsets**
- **Annotation verschiedener Spuren (zur Beantwortung unterschiedlicher Fragestellungen)**
- **Entscheidung über maschinelles / manuelle**

- EXMARaLDA als Grundlage für Architektur?
- Spuren für:
 - Texte
 - POS-Tagging
 - Zielhypothese(n) – Problem gerade in unteren Klassen viele Fehler
 - ...
- Metainformationen
- Testergebnisse separat ablegen (derzeit SPSS)

15

- Becker-Mrotzek, M. & Böttcher, I. (2006). Schreibkompetenz entwickeln und beurteilen. Cornelsen Scriptor, Berlin.
- Biber, D. (2009). In A. Lüdeling & M. Kytö (Hrsg.), *Corpus Linguistics. An International Handbook*. Mouton de Gruyter, Berlin.
- Bird, S. & Simons, G. (2003). Seven Dimensions of Portability for Language Documentation and Description. In *Language* 79, 557-582.
- Bennöhr, J. (2007). A Web-based Personalised Textfinder for Language Learners. In G. Rehm, A. Witt, L. Lemnitzer (Hrsg.), *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen - Proceedings der GLDV-Jahrestagung 2007, Tübingen*. Gunter Narr Verlag, Tübingen, pp. 225-234.
- Bredel, U. & Reich, H. H. (2008). Literale Basisqualifikationen I und II. In K. Ehlich, U. Bredel, H. H. Reich (Hrsg.), *Bildungsforschung Band 29/I: Referenzrahmen zur altersspezifischen Sprachaneignung*. Bundesministerium für Bildung und Forschung (BMBF), Berlin, Bonn.
- Büchner, I. & Bennöhr, J. (2007). Unveröffentlichte Weiterbearbeitung des Hamburger Leitfadens für die Bewertung von Textprodukten.
- Burstein, J., Chodorow, M. & Leacock, C. (2003). CriterionSM Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. *IAAI 2003*: 3-10.

16

- Ehlich, K., Bredel, U. & Reich, H. H. (2008). Sprachaneignung – Prozesse und Modelle. In K. Ehlich, U. Bredel, H. H. Reich (Hrsg.), *Bildungsforschung Band 29/I: Referenzrahmen zur altersspezifischen Sprachaneignung*. BMBF, Berlin, Bonn.
- Grießhaber, W. (2006). Die Entwicklung der Grammatik in Texten vom 1. bis zum 4. Schuljahr. In B. Ahrenholz (Hrsg.), *Kinder mit Migrationshintergrund - Spracherwerb und Förder-möglichkeiten*. Freiburg i. B., Fillibach, 150-167.
- Jude, N. & Wirth, J. (2007). Neue Chancen bei der technologiebasierten Erfassung von Kompetenzen. In J. Hartig und E. Klieme (Hrsg.), *Bildungsforschung Band 20: Möglich-keiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*, Bundesminis-terium für Bildung und Forschung, Berlin, Bonn.
- Kakkonen, T., Myller, N., & Sutinen, E. (2004). Semi-Automatic Evaluation Features in Computer - Assisted Essay Assessment. *Proceedings of the 7th IASTED Conference on Computers and Advanced Technology in Education (CATE 2004)*, pp. 456-461. Kauai, Hawaii, USA.
- Lenhard, W., Baier, H., Hoffmann, J. & Schneider, W. (2007). Automatische Bewertung offener Antworten mittels Latenter Semantischer Analyse. *Diagnostica*, 53, 155-165.
- Liu, X., Croft, W. B., Oh, P. & Hart, D. (2004). Automatic Recognition of Reading Levels from User Queries. In *Annual Proceedings of the 27th annual international conference on Re-search and development in information retrieval*, Sheffield, United Kingdom. pp. 548-549.
- Lüdeling, A. (2008). Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In M. Walter & P. Grommes, Patrick (Hrsg.), *Fortgeschrittene Lerner-varietäten*. Niemeyer, Tübingen, pp. 119-140.
- May, P. & Juchems, A. (1996). Hamburger Leitfaden für die Bewertung von Bild- und Textprodukten (**HLBT**) zur Erfassung bildnerischer und textueller Gestaltungsfähigkeiten von Kindern. Universität Hamburg: Typoskript.

- McNamara, D. S. & the CSEP lab (2003). Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Annual project report submitted to the Institute of Education Sciences.
- Nussbaumer, M. & Sieber, P. (1994). Texte analysieren mit dem Zürcher Textanalyseraster. In: Sieber, P. (Hrsg.) *Sprachfähigkeiten - Besser als ihr Ruf und nötiger denn je!* Aarau u. a., 141-186.
- Quinlan, T., Higgins, D. & Wolff, S. (2009). Evaluating the Construct-Coverage of the e-rater® Scoring Engine, ETS, Princeton, New Jersey.
- Röber, C. (2007). Schriftlernen unter Berücksichtigung des kindlichen Schriftwissens. In H. Schöler & A. Welling (Hrsg.), *Sonderpädagogik der Sprache*. Göttingen, Bern u.a.: Hogrefe, S. 148-172.
- Si, L. & Callan, J. (2001). A Statistical Model for Scientific Readability. In *Proceedings of the tenth international conference on Information and knowledge management*, Atlanta, Georgia, USA. pp. 574-576.
- Uitdenbogerd, Alexandra L. (2003). Using the Web as a Source of Graded Reading Material for Language Acquisition. In *Lecture Notes in Computer Science*, pp. 423-432, Vol. 2783.
- Valenti, S., Neri, F. & Cucchiarelli, A. (2003). An Overview of Current Research on Automated Essay Grading. In *Journal of Information Technology Education*, pp. 220-230, Vol. 2.
- Williams, R. & Dreher, H. (2004). Automatically grading essays with markit? Issues in Informing Science and Information Technology 1: 0693-0700.